A perspective on data quality

ESP Solutions Group

Glynn D. Ligon, Ph.D.

CEO, ESP Solutions Group

1101000101000101001111101010111010100101001010100100101001010110
1010111010001010001010011111010101110101001010010101001001010010
1011010101110100010100010100111110101011101010010100101010010010
1001010110101011101000101000101001111101010111010100101001010100
1001010010101101010111010001010001010011111010101110101001010010
1010010010100101011010101110100010100010100111110101011101010010
1001010100100101001010110101011101000101000101001111101010111010
1001010010101001001010010101101010111010001010001010011111010101
1101010010100101010010010100101011010101110100010100010100111110
1010111010100101001010100100101001010110101011101000101000101001
1111010101110101001010010100100101001010110101011101000101000101
0100111110101011101010010100101010010010100101011010101110100010
1000101001111101010111010100101001010100100101001010110101010101

ESP Solutions Group

Data Quality, *the basics*

1. Get data right from the start.

2. Keep them right at every step.

"Every 1" creates and passes on only perfect data. Yes, you're the 1.

3. Give people help to do this.

The next person in line can't fix the last person's errors as easily as that person can.

ESP Solutions Group

Every 1 Matters: Missing Data

1101000101000101001111101010111010100101001010100100101001010110
1010111010001010001010011111010101110101001010010101001001010010
1011010101110100010100010100111110101011101010010100101010010010
1001010110101011101000101000101001111101010111010100101001010100
1001010010101101010111010001010001010011111010101110101001010010
1010010010100101011010101110100010100010100111110101011101010010
1001010100100101001010110101011101000101000101001111101010111010
1001010010101001001010010101101010111010001010001010011111010101
1101010010100101010010010100101011010101110100010100010100111110
1010111010100101001010100100101001010110101011101000101000101001
1111010101110101001010010101001001010010101101010111010001010001
0100111110101011101010010100101001010010101101010111010001010001
1000101001111101010111010100101001010100100101001010110101010101

Every 1 Matters: Bad Data

1101000101000101001111101010111010100101001010100100101001010110
1010111010000101000101001111101010111010100101001010100100101001010010
1011010101110100010100010100111110101011101010010100101010010010
1001010110101011101000101000101001111101011101010010100101010100
1001010010101101010111010001010001010011111010101110101001010010
1010010010100101010110101011101000101000101001111101010111010100101001010
1001010100100101001010110101011101000101000101001111101010111010
1001010010101001001010010101110101011101000101000101001111101010101
1101010010100101010010010100101011010101011101000101000101001111101
1010111010100101001010100100101001010110101010111010001010001010011
1111010101110101001010010101001001010010101101010111010001010001
0100111111010101110101001010010100100101001010110101011101001000101
1000101001111101010111010100101001010100100101001010110101010101

ESP Solutions Group

Every 1 Matters: Late Data

110

Every 1 Matters: Quality Data

ESP Solutions Group

Every 1 Matters

Student Activity

Every 1 Matters: Quality Data

ESP Solutions Group

The Quintessential Student Record

For a school, the quintessential student record is the physical or virtual cumulative, longitudinal data within the school's student information system.

This record is the authoritative source for:

- **EASIER**
- **State Reporting**
- **Federal Reporting**
- **Official Transcripts**
- **Registration**
- **Scheduling**
- **Grade Reporting**
- **Attendance Accounting**

ESP Solutions Group

The Essence of Error-Free Exchange

Most vulnerable times for data:

- **Entry**
- **Exchange**

Similar to take offs and landings for airplanes

ESP Solutions Group

The Truths of Data Quality

The Four Great Truths about Data Quality: Data quality is highest when…

1. The data providers know what's expected.
2. The data providers use the data themselves for their own work.
3. Everyone, everywhere checks the data.
4. The data are available and used.

ESP Solutions Group

The Principles of Data Quality

The *Expectation* Principle of Data Quality

- Data quality can only be achieved when the expectations are clear.

ESP Solutions Group

The Principles of Data Quality

The *Use* Principle of Data Quality

- **Data quality matters when the data are used by the person collecting and reporting the data.**

ESP Solutions Group

The Principles of Data Quality

The *Check Your Work* Principle of Data Quality

- Data quality requires “every 1” to check their own data.

ESP Solutions Group

The Principles of Data Quality

The *Comparability* Principle of Data Quality

- **Data quality matters when the data are compared.**

ESP Solutions Group

The Principles of Data Quality

The *Hierarchical Norm* Principle of Data Quality

Institutional Hierarchy

US Department of Education

State Education Agency

Local Education Agency (District)

School

Employee

The Principles of Data Quality

The *Hierarchical Norm* Principle of Data Quality

Institutional Hierarchy

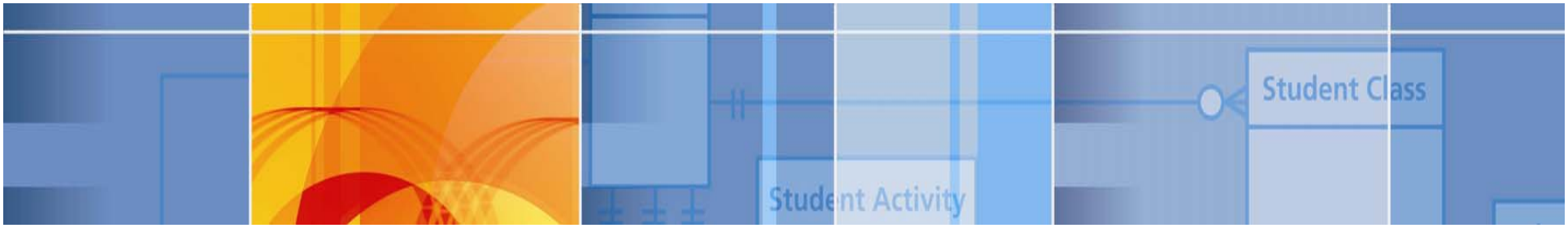
US Department of Education

State Education Agency

Local Education Agency (District)

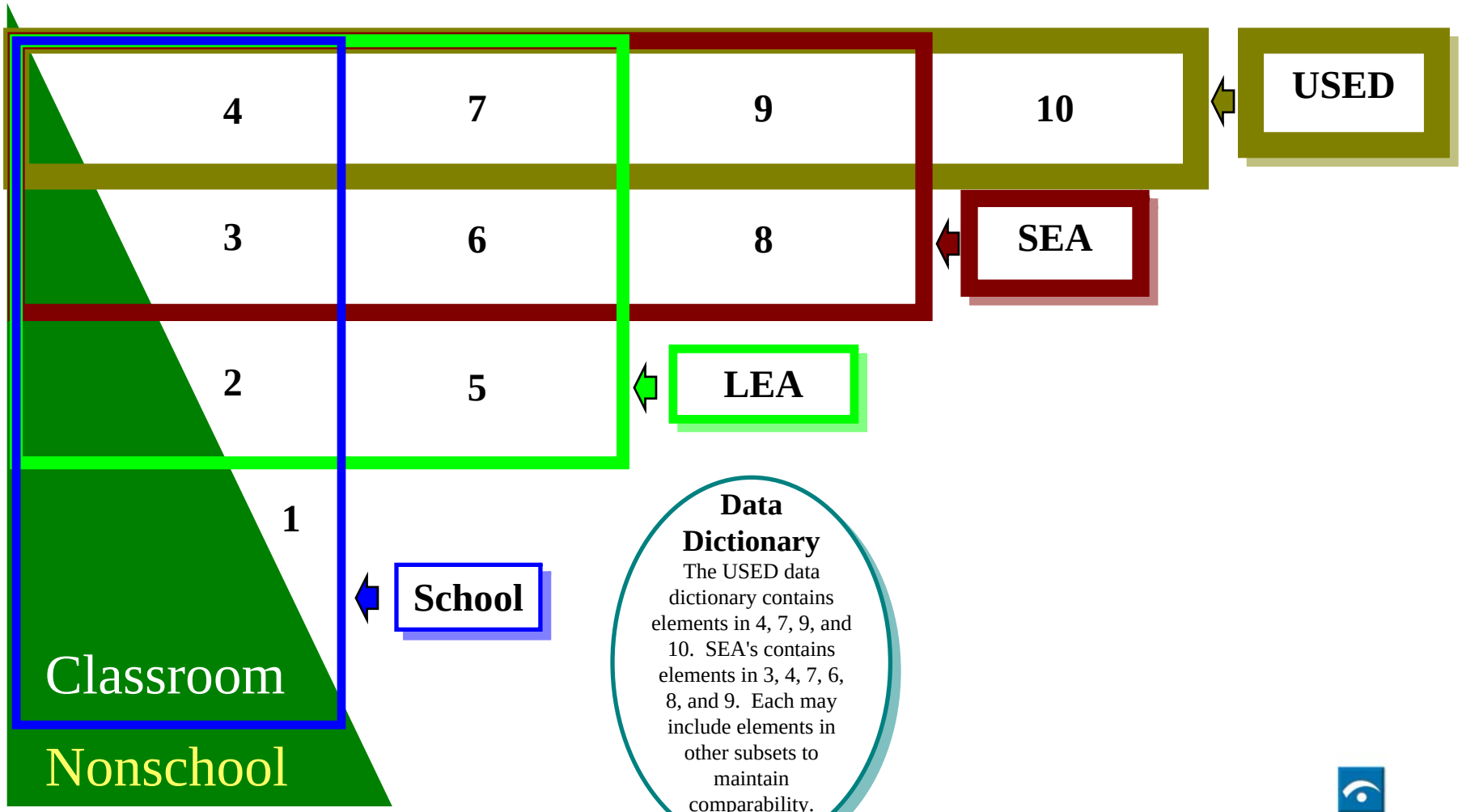
School

Employee



Data Sharing across Levels of the Education Enterprise

Rectangular Venn Diagram for Sharing Data Across Education Agencies



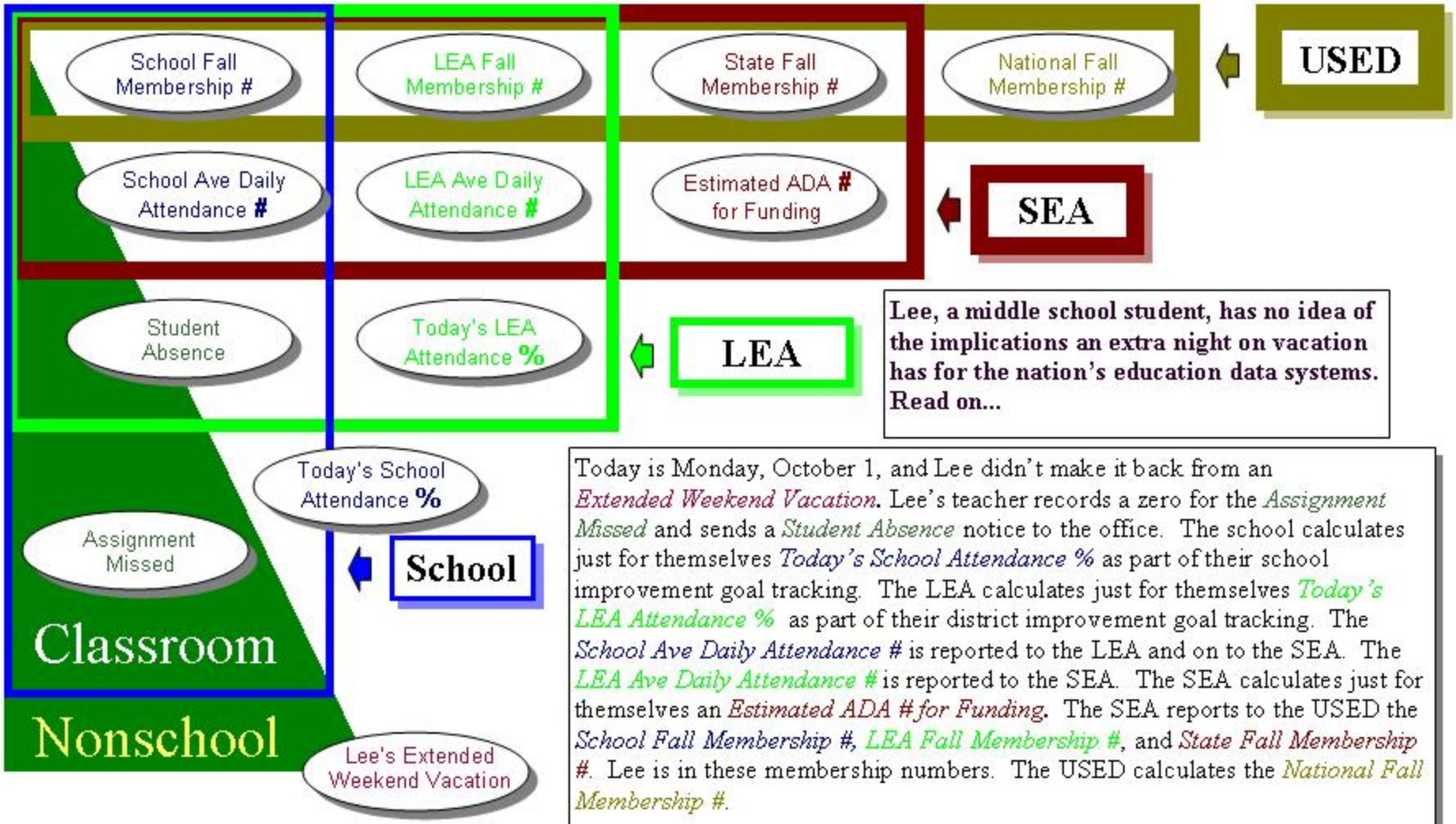
Data Dictionary

The USED data dictionary contains elements in 4, 7, 9, and 10. SEA's contains elements in 3, 4, 7, 6, 8, and 9. Each may include elements in other subsets to maintain comparability.

ESP Solutions Group

How Data are Shared

Rectangular Venn Diagram for Sharing Data Across Education Agencies



Lee, a middle school student, has no idea of the implications an extra night on vacation has for the nation's education data systems. Read on...

Today is Monday, October 1, and Lee didn't make it back from an *Extended Weekend Vacation*. Lee's teacher records a zero for the *Assignment Missed* and sends a *Student Absence* notice to the office. The school calculates just for themselves *Today's School Attendance %* as part of their school improvement goal tracking. The LEA calculates just for themselves *Today's LEA Attendance %* as part of their district improvement goal tracking. The *School Ave Daily Attendance #* is reported to the LEA and on to the SEA. The *LEA Ave Daily Attendance #* is reported to the SEA. The SEA calculates just for themselves an *Estimated ADA # for Funding*. The SEA reports to the USED the *School Fall Membership #*, *LEA Fall Membership #*, and *State Fall Membership #*. Lee is in these membership numbers. The USED calculates the *National Fall Membership #*.

The *Transformation* Principle of Data Quality

- **A subject institution may define a data element differently from its authority only to the extent that the data element can be derived from or transformed into the precise definition of the higher authority.**

The Principles of Data Quality

The *Transformation Burden* Principle of Data Quality

- **Part A: The burden to transform is solely the burden of the subject institution.**
- **Part B: This burden compels the subject institution to comply with the standard of the authority.**

ESP Solutions Group

The Principles of Data Quality

The *Monkey on My Back Version of the* Transformation Burden Principle of Data Quality

- **Data Provider: I can get them to clean the data because they are the ones who want it anyway.**
- **Data Requestor: I'm the one who needs these data, so I have to clean them up if they won't.**

ESP Solutions Group

The Principles of Data Quality

The *Invented Here* Principle of Data Quality

- **Competes with the Transformation Burden Principle**
- **As the local expert, I know how we should define our data.**

ESP Solutions Group

The Principles of Data Quality

The *Vendor Rules* Principle of Data Quality

- **When we chose our vendor, we chose our data standards.**

ESP Solutions Group

The *Inertia* Principle of Data Quality

- **If we change to use the authority's standard, we have to retrain everyone and reconfigure all our software.**

The Principles of Data Quality

What does this mean for me?

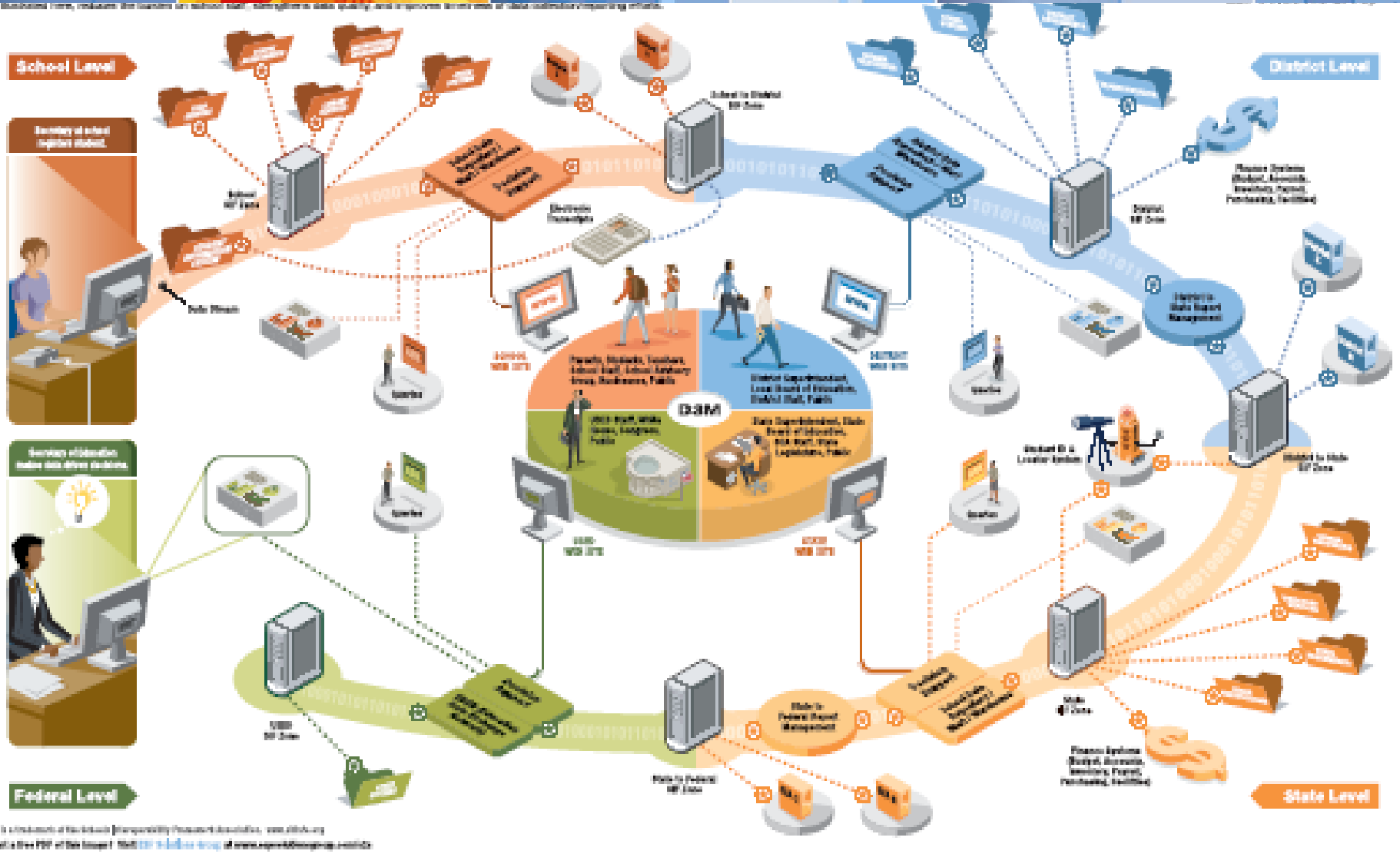
- If you follow the authority's rules, burden is lower.
- If you change the rules, you have to re-work your data for reporting.

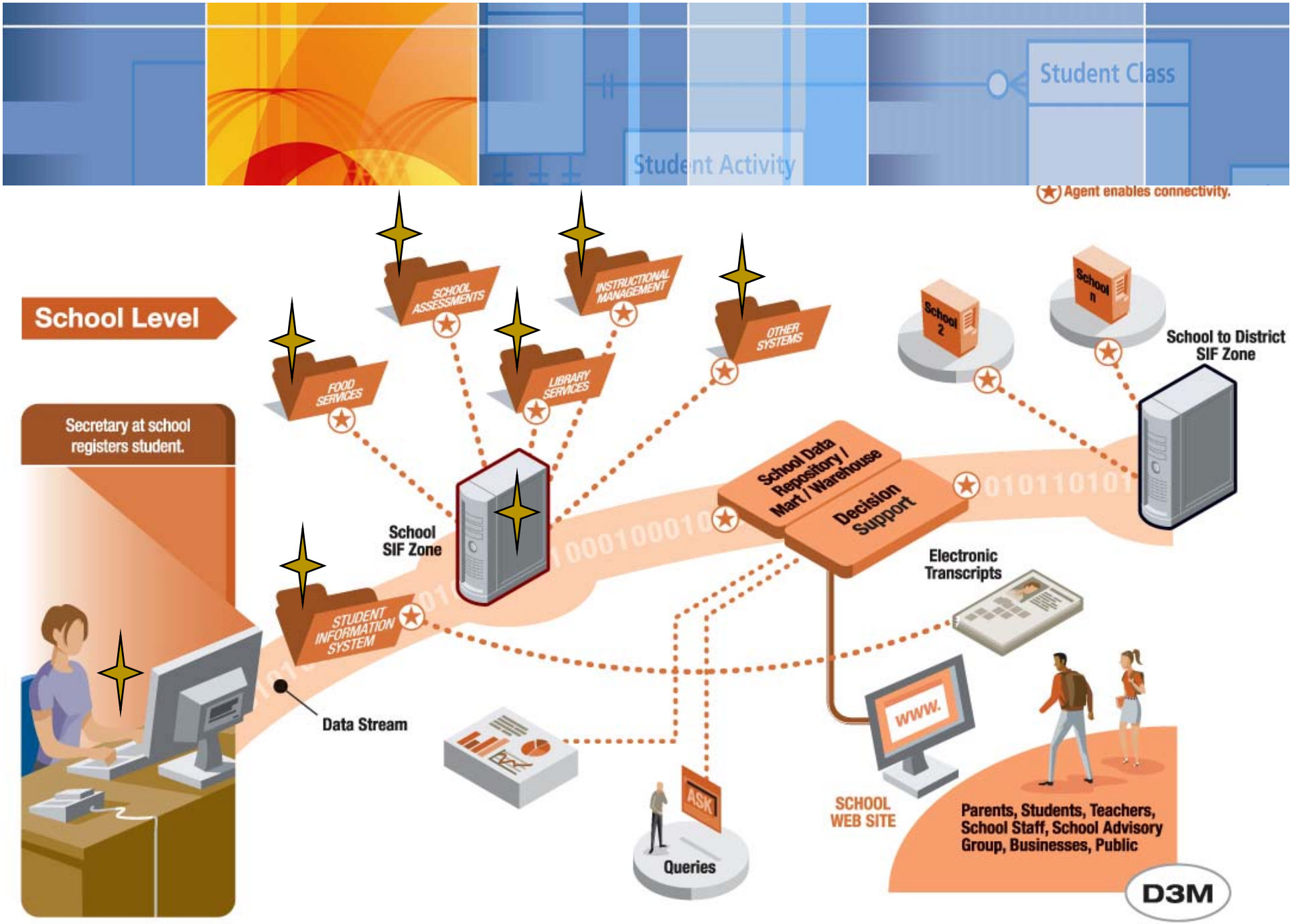
What does this mean for data quality?

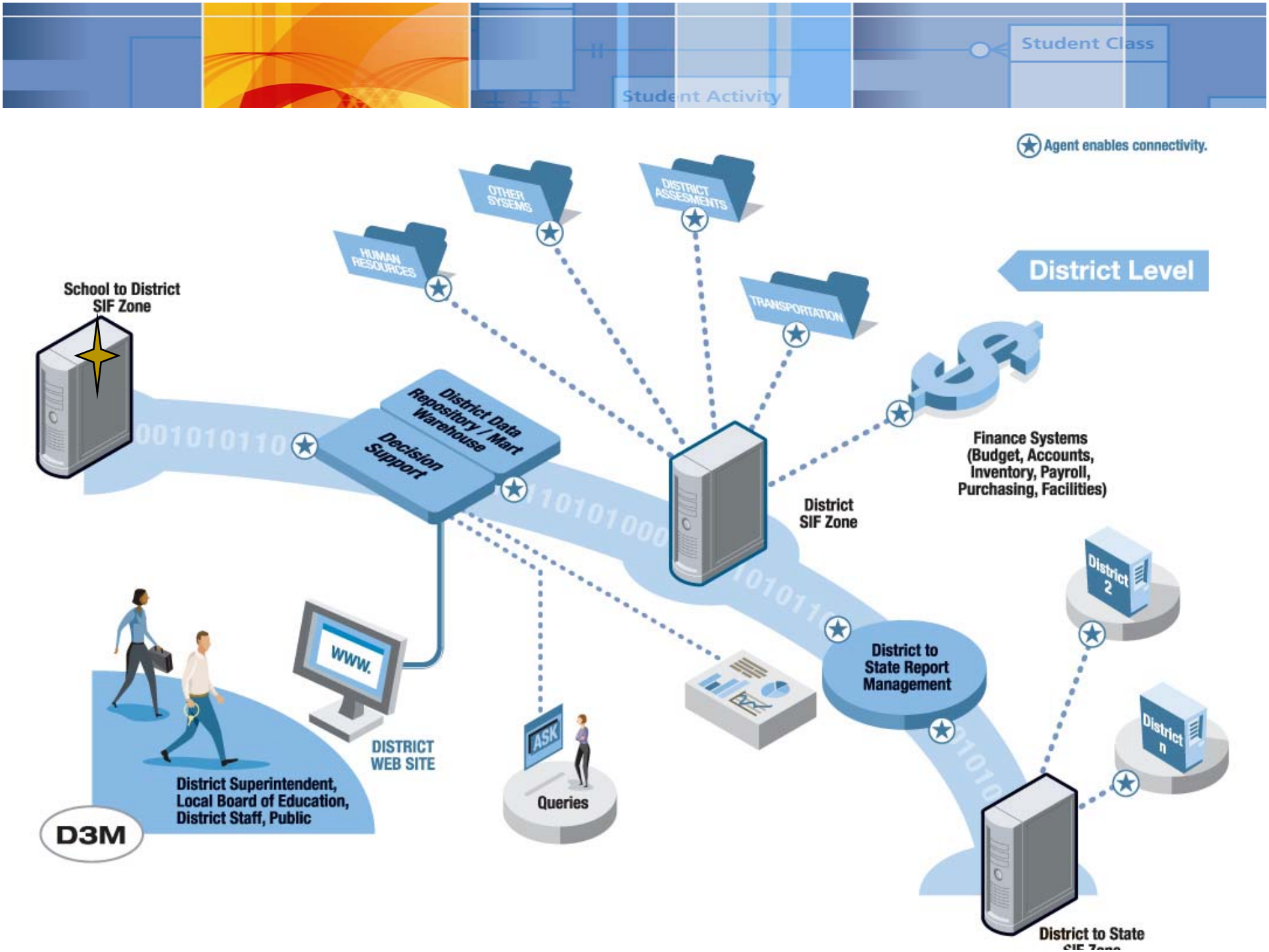
- If people follow the rules, quality is higher.
- If people change the rules, quality is tougher to achieve.

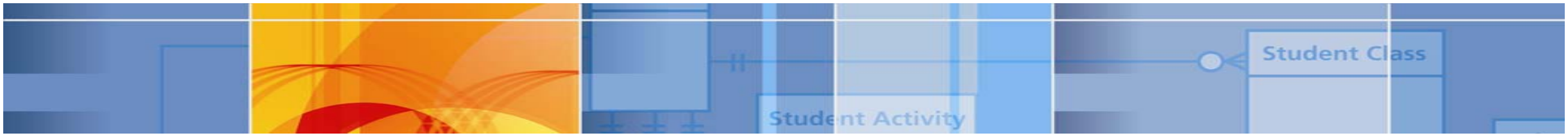
ESP Solutions Group

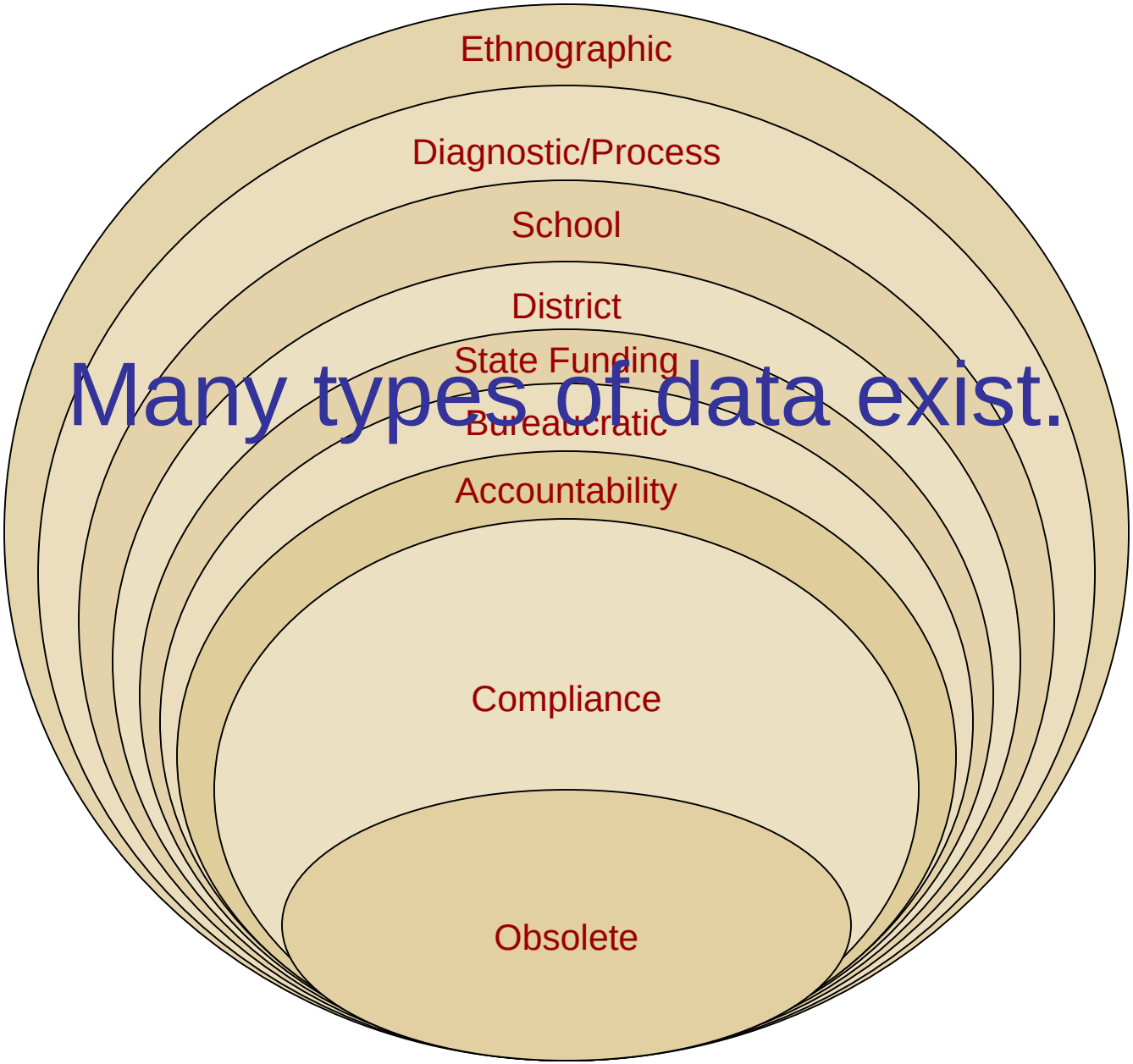
Secretary to Secretary




Student Class
Student Activity
Agent enables connectivity.
School Level
SCHOOL ASSESSMENTS
INSTRUCTIONAL MANAGEMENT
OTHER SYSTEMS
FOOD SERVICES
LIBRARY SERVICES
School 2
School n
School to District SIF Zone
Secretary at school registers student.
School SIF Zone
School Data Repository / Mart / Warehouse
Decision Support
Electronic Transcripts
STUDENT INFORMATION SYSTEM
Data Stream
www.
SCHOOL WEB SITE
ASK
Queries
Parents, Students, Teachers, School Staff, School Advisory Group, Businesses, Public
D3M

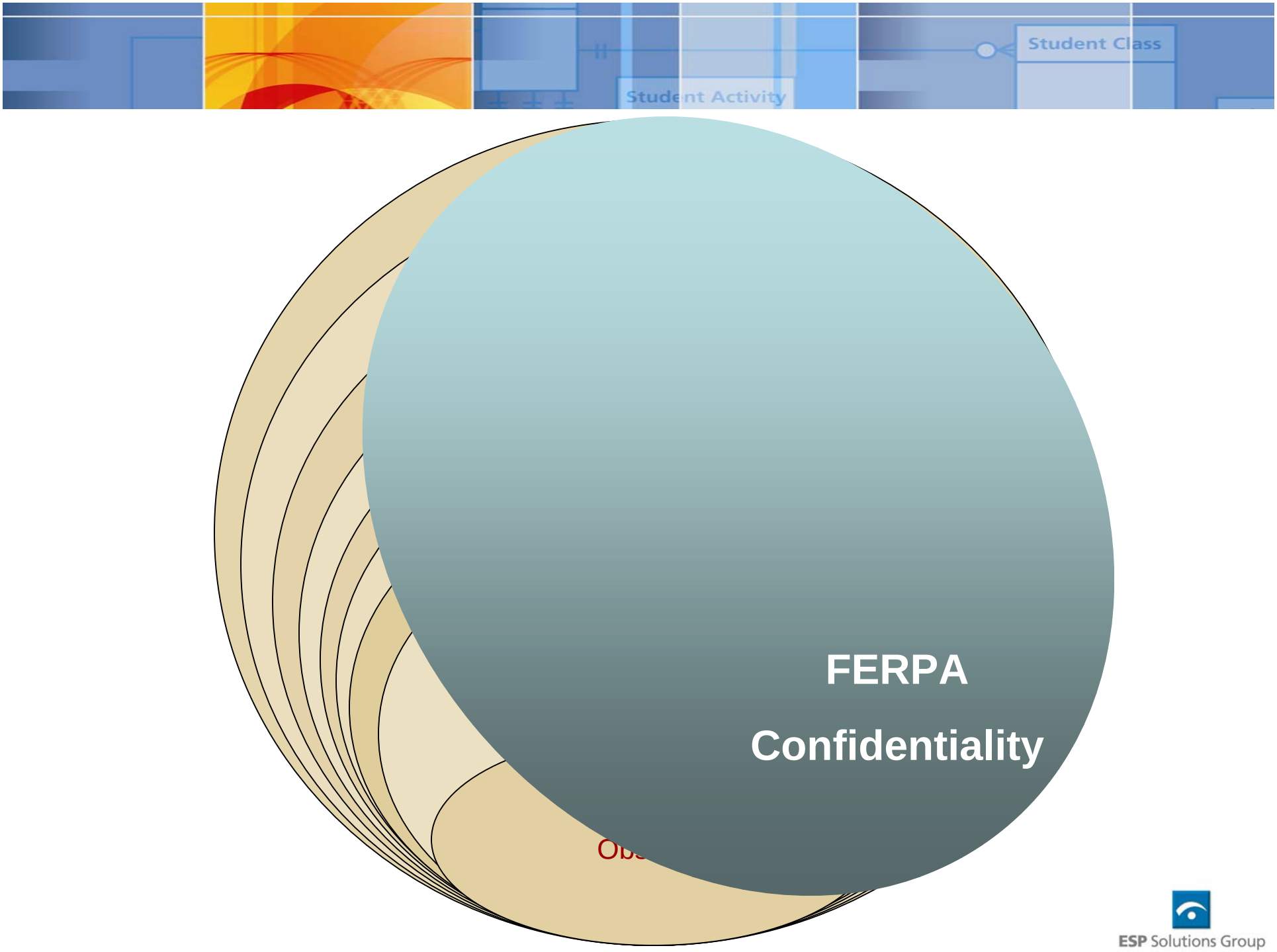

Student Class
Student Activity
Agent enables connectivity.
District Level
OTHER SYSEMS
DISTRICT ASSESSMENTS
HUMAN RESOURCES
TRANSPORTATION
School to District SIF Zone
District Data Repository / Mart Warehouse
Decision Support
Finance Systems (Budget, Accounts, Inventory, Payroll, Purchasing, Facilities)
District SIF Zone
District 2
District to State Report Management
District n
www.
DISTRICT WEB SITE
ASK
Queries
District Superintendent, Local Board of Education, District Staff, Public
D3M
District to State



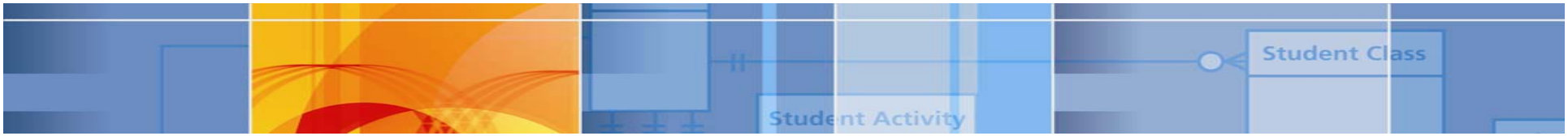


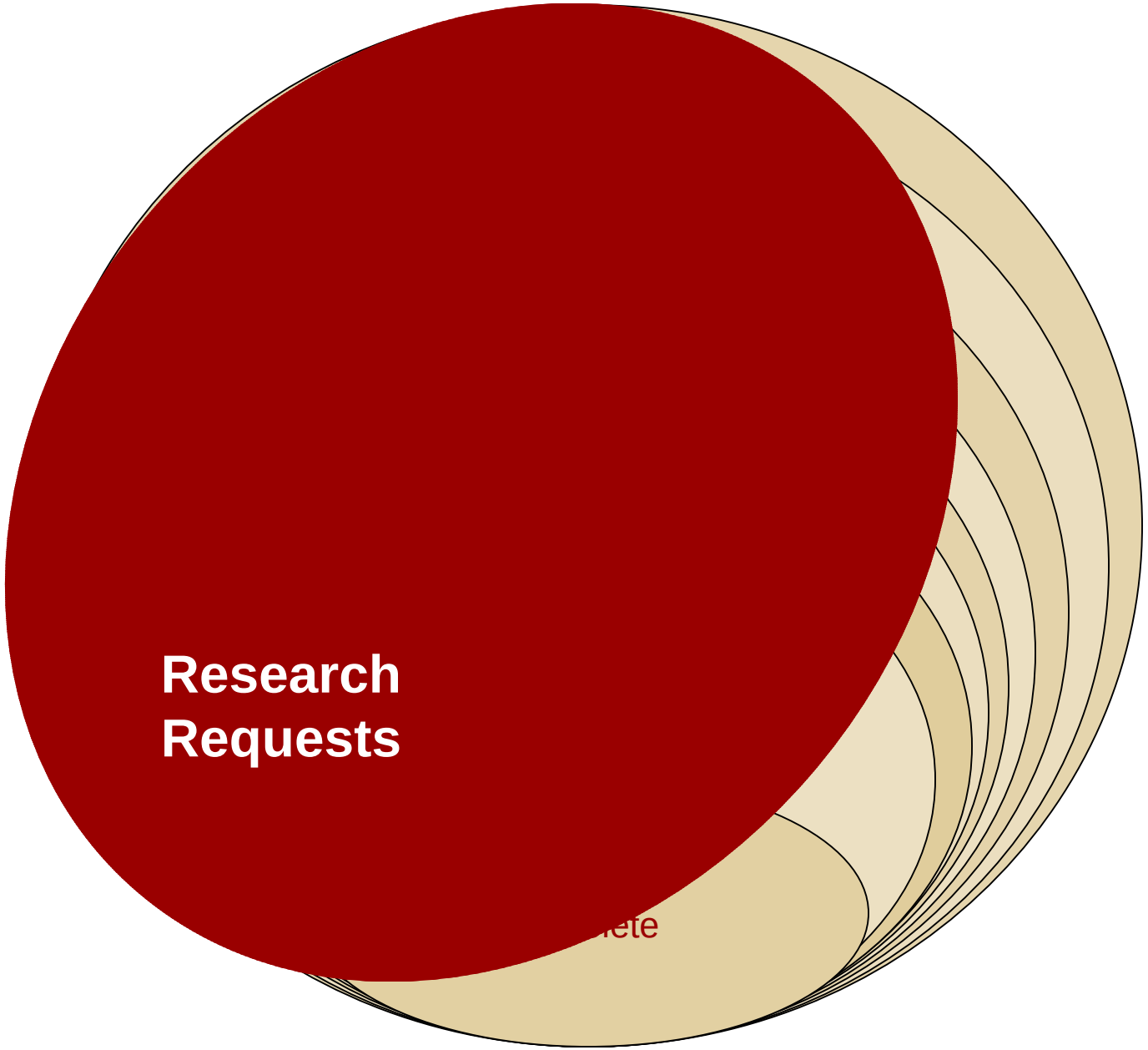
Many types of data exist.

ESP Solutions Group

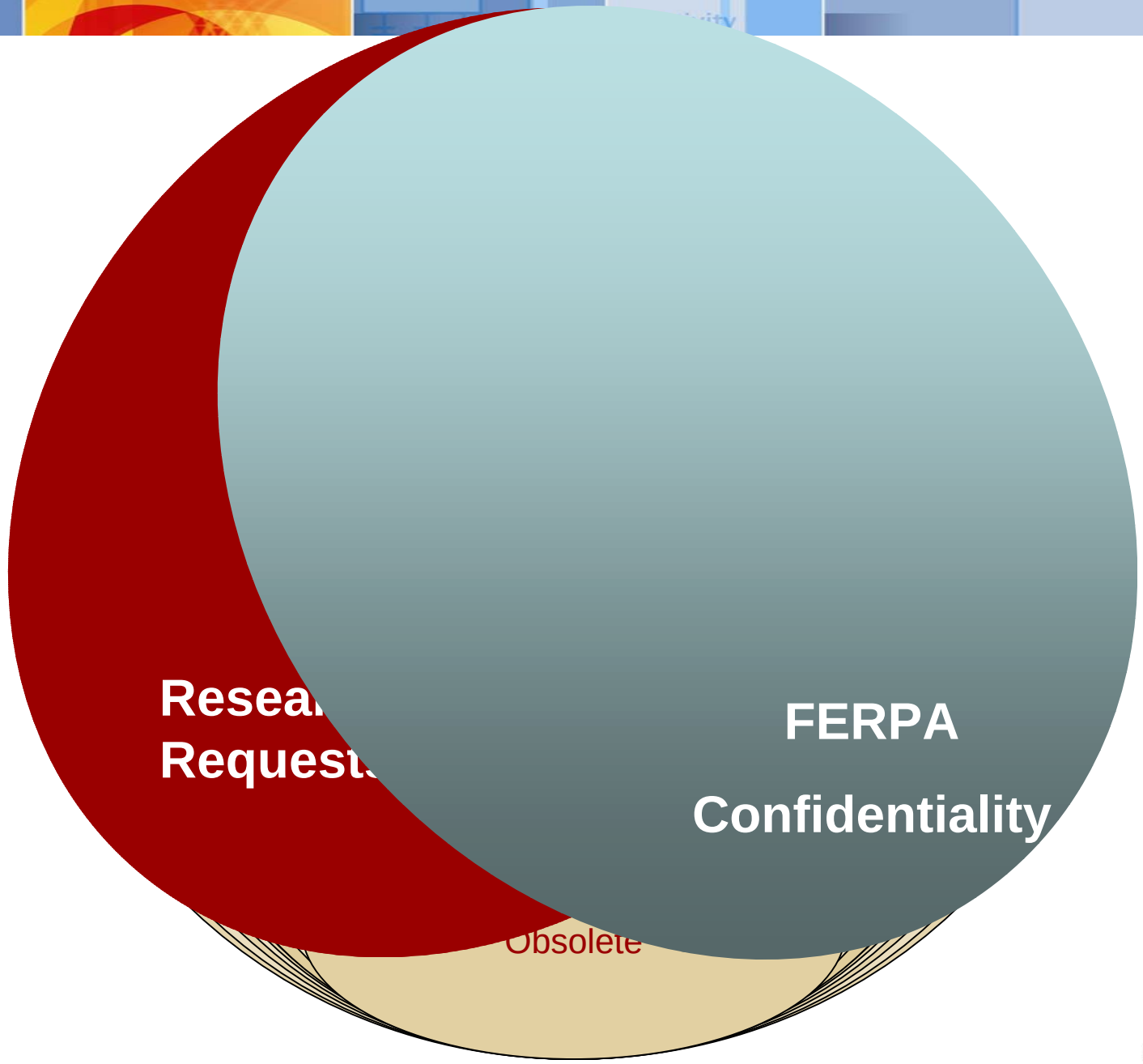


ESP Solutions Group





ESP Solutions Group

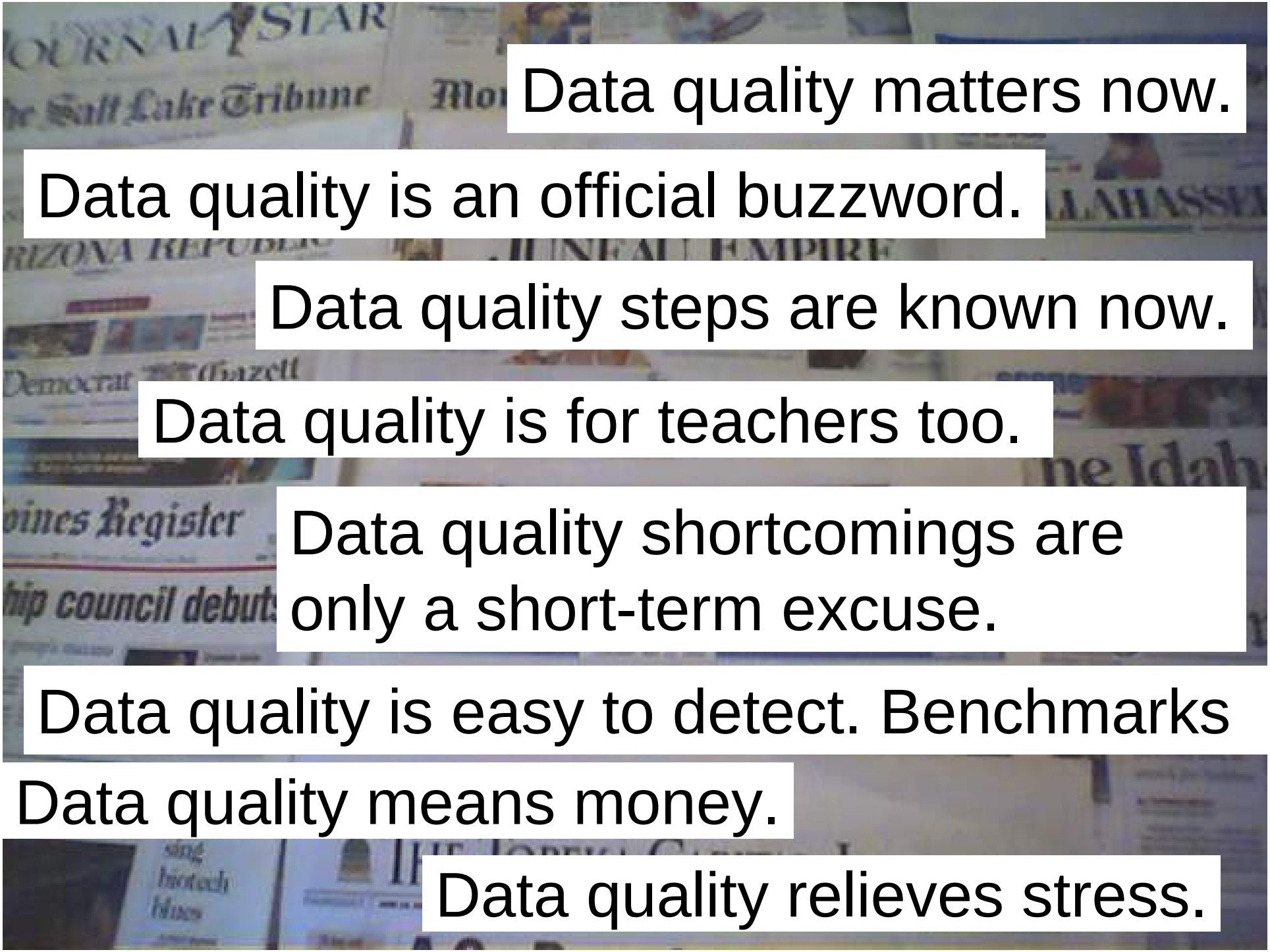


ESP Solutions Group

News from Around the Nation

What do we know about data quality from our work around the nation?

ESP Solutions Group

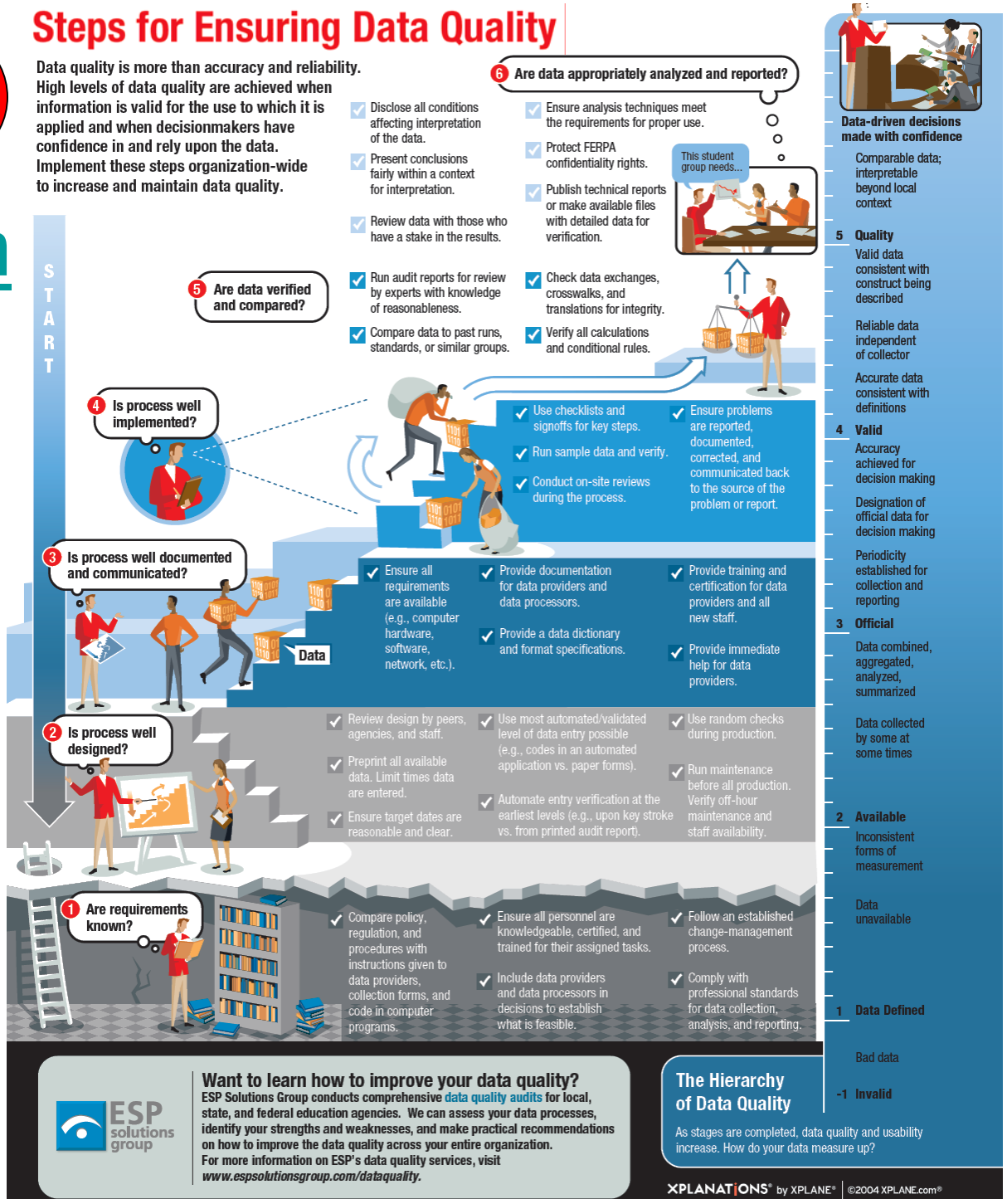

Data quality matters now.
Data quality is an official buzzword.
Data quality steps are known now.
Data quality is for teachers too.
Data quality shortcomings are only a short-term excuse.
Data quality is easy to detect. Benchmarks
Data quality means money.
Data quality relieves stress.

6

www.espsg.com

Sign up for…

My ESP Page



Disclaimer: This 6-step program is not associated with any other 12-step or 7-step or other numbered program.

However, before we begin, it would be helpful for each of us to acknowledge that we have abused data in the past.

Each morning when we get to work, we will pause and reflect that we have gone another day without providing any bad data to anyone else.

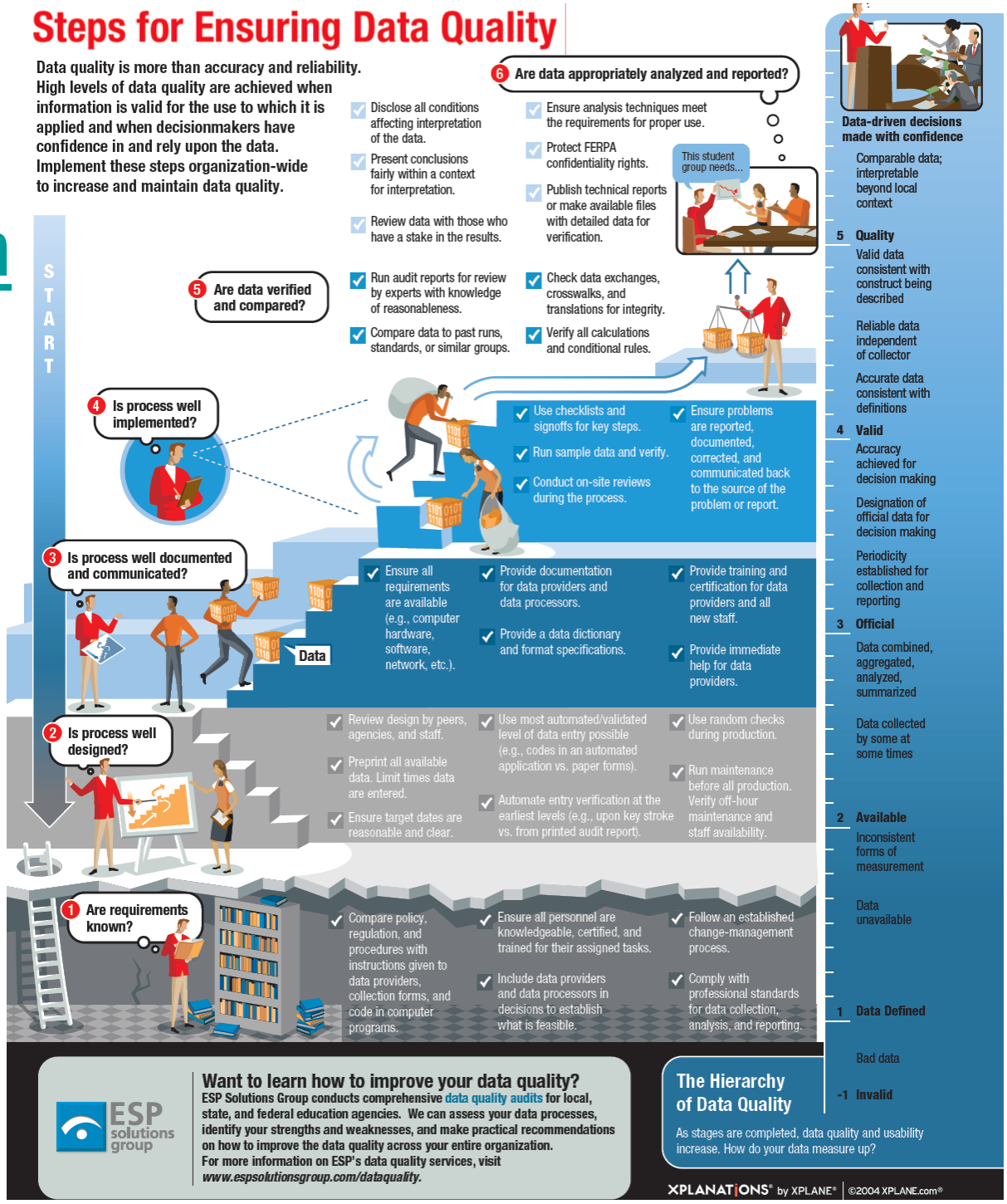
We trust in a higher power, the Iowa Department of Education, to set the standards for our data.

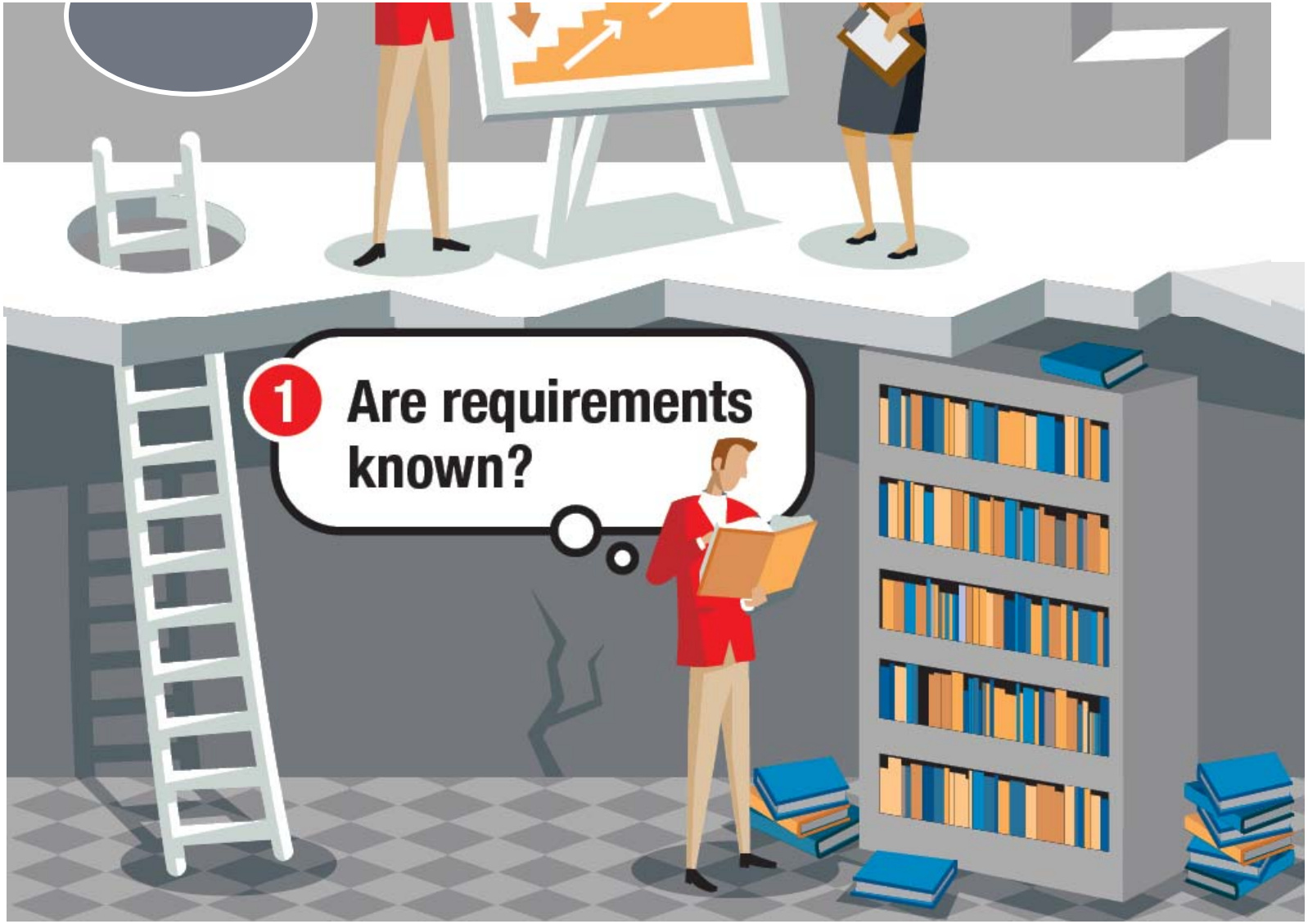
Our goal is to provide quality--one datum at a time.

www.espsg.com

Sign up for…

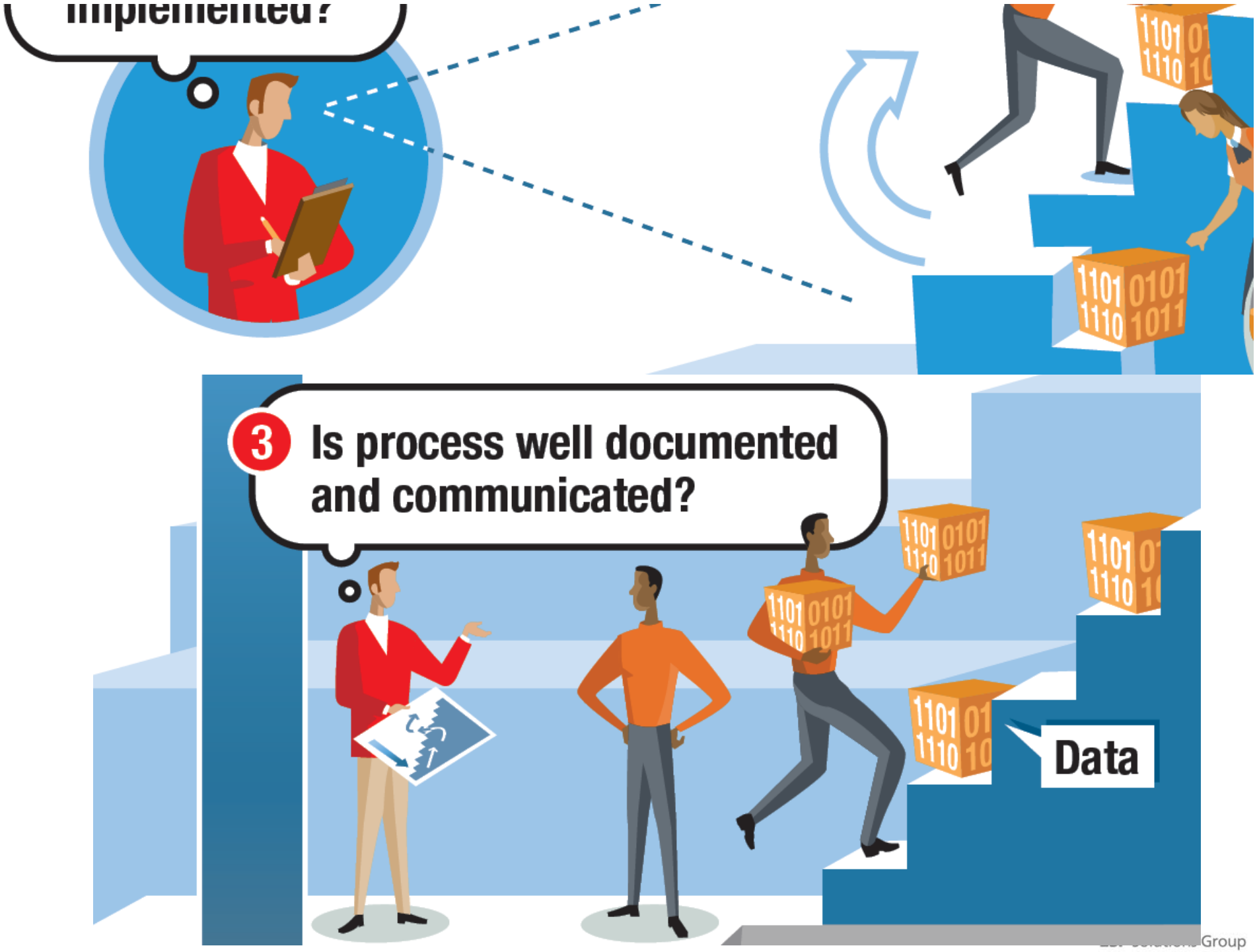
My ESP Page

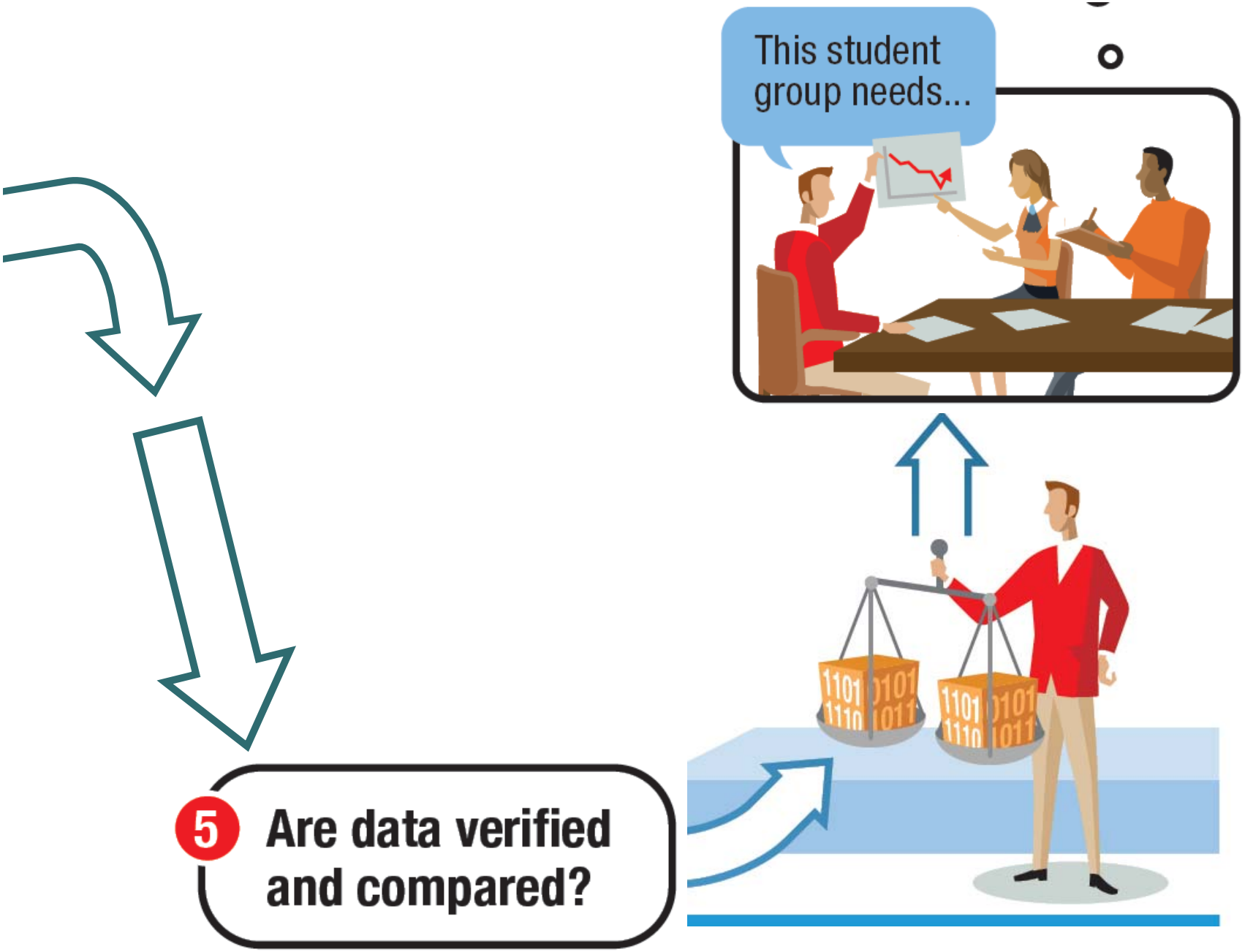




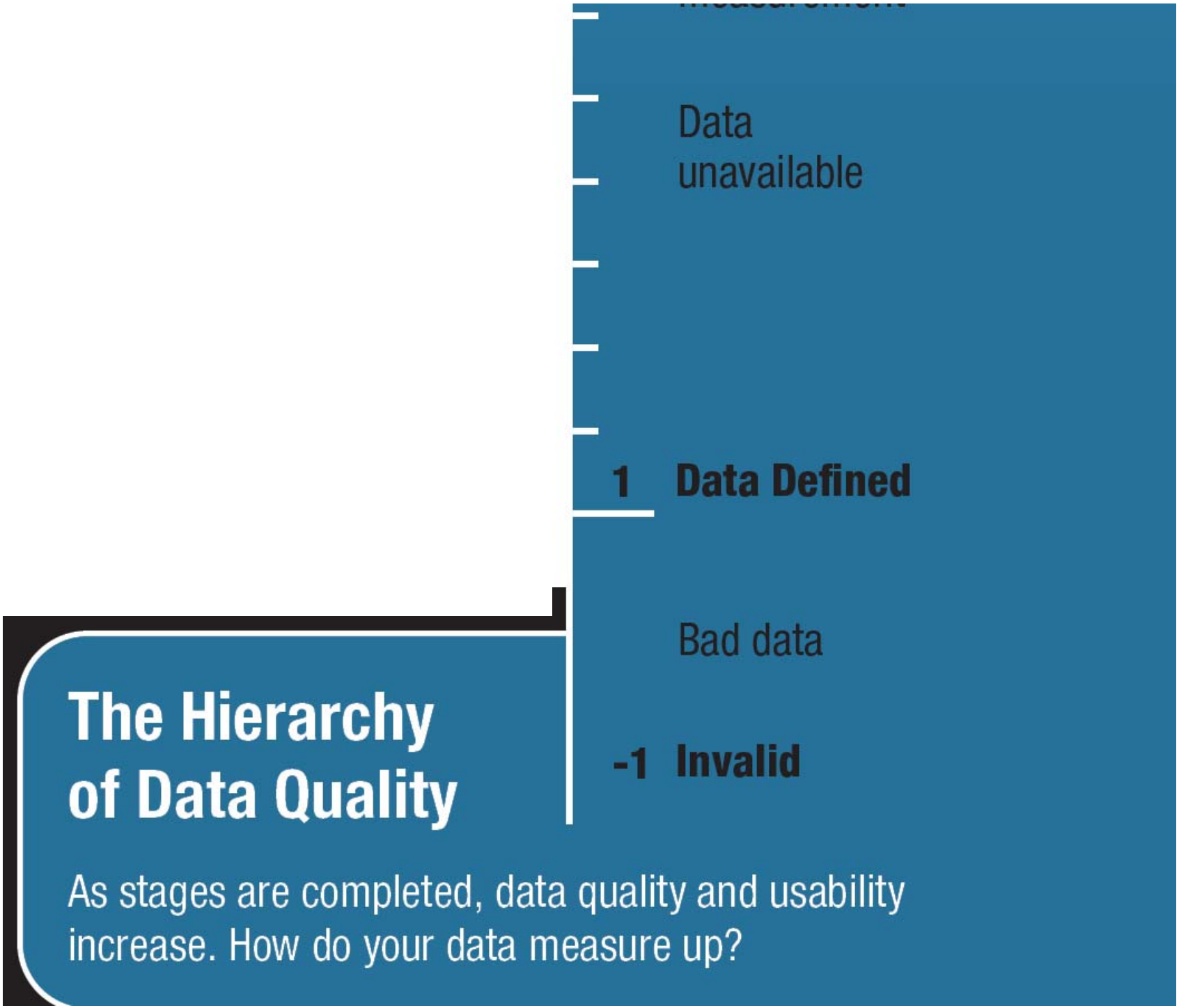
Steps for Ensuring Data Quality

ESP Solutions Group


implemented?
3 Is process well documented and communicated?
Data



ESP Solutions Group


Data
unavailable
1 Data Defined
Bad data
-1 Invalid
The Hierarchy
of Data Quality
As stages are completed, data quality and usability
increase. How do your data measure up?

As we review these grievous errors that other people have made with their data, please raise your hand if you know of someone (not yourself, of course, but a coworker, a neighbor, or a superintendent) who has committed one of these data transgressions.

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be lackadaisical when the requirements are precise

arg

ic

r

gi
sa
th

nts
were born
esterday

be

abbrev

d enough

ke

nstalled

call

co

thir

ght on
the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

make notes in data fields

First Name Field:
"Mandy (but mother says she prefers to be called "Pookey")"

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

arg

gi
sa
th

abbreviate to save t

keep doing thi

call up a friend a

copy and paste

copy and paste from one file (format) to another

ic

nts
born
day

re is good enough

are was installed

data

Pat	M	Johnson	Jr
Johnson, Pat M, Jr			

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be lackadaisical when the requirements are precise

ic

gi
sa
th

nts
born
day

abbreviate to sa

re is good enough

keep doing thi

are was installed

call up a frie

data

copy

Required

Patrick	M.	Johnson	Jr.
Pat		Johnson	

Lackadaisical

of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

gi
sa
the field

ask for forgiveness rather than permission

students were born yesterday

be creativ

abbreviate to s

e is good enough

keep doin

re was installed

call up a fri

ata

copy and

think of dat

ything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

add codes to be more specific

1 = Graduate

2 = Transfer

3 = Retainee

U = Unknown

M = Sent to Marie for Coding

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be lackadaisical when the requirements are precise

make the data your own

be right when the world is wrong

give everyone the same value just to fill the field

practice civil disobedience

ask for forgiveness rather than permission

say students ere born esterday

be

abbrevi

l enough

ke

stalled

call

copy and paste without being extra careful

Phone Number Field:
"555-555-5678 except on Tue then 656-555-5555"

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

gi
sa
the field

ask for forgiveness rather than permission

students
ere born
esterday

be

abbrevi

l enough

ke

stalled

give everyone the same value just to fill the field

SSN Field:
"111-11-1111"

call up a friend at IDE and ask for her/him to correct your data

copy and paste without being extra careful

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

gi
sa
the field

ask for forgiveness rather than permission

students
ere born
esterday

be

abbrevi

l enough

ke

stalled

call

co

thin
the

ht on

pass data entry on to someone who doesn't know the rules or can't follow them

ask for forgiveness rather than permission

"Oh, hello, yeah, I think I may have accidentally left all the Title 1 codes off my file. I'm really sorry. Can you ever forgive me?"

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

argue with official names, spelling, or capitalization

gi
sa
th

ic

nts
born
day

abbrevi

l enough

ke

stalled

call

co

thin
the

ght on

pass
them

ollow

District Name Field:
"Colorado Springs"
(Official Name: El Paso County District 11)

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be lackadaisical when the requirements are precise

be right when the world is wrong

gi… sa… the field

ask for forgiveness rather than permission

students were born yesterday

be creative to get double use from the data

abbreviate to save time

be better when the software is good enough

keep doing things the way you did before the new software was installed

call up a friend at IDE and ask for her/him to correct your data

copy and paste without being extra careful

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

gi
sa
the field

ask for forgiveness rather than permission

students
ere born
esterday

be

abbrevi

l enough

ke

stalled

say students were born yesterday

Birth Date Field:
"April 11, 2006"

call up a friend at IDE and ask for her/him to correct your data

copy and paste without being extra careful

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

be creative to get double use from the data

gi
sa
th

ic

nts
born
day

abbreviate to save time

be better when the software is good enough

ke

stalled

call

co

Course Field:
"Lunch A"

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be lackadaisical when the requirements are precise

be better when the software is good enough

abbreviate to save time

be better when the software is good enough

call

Gender Field:
"Female"
(Valid Code = F)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

r

gi
sa
th

nts
born
day

ab

ough

ed

keep doing things the way you did before the new software was installed

call up a friend at IDE and ask for her/him to correct your data

co

thin
the

ght on

pass
them

ollow

"My Excel spreadsheet is really the official record for my students."

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

call up a friend at IDE and ask for her/him to correct your data

gi
sa
th

ic

nts
born
day

abbreviate to save time

be better when the software is good enough

ke

stalled

call

co

"Hi Coleen, would you be a dear and just change those LEP codes for me again this year?"

thin
the

ght on

pass

ollow

them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

gi
sa
the field

on

students were born yesterday

be creativ

abbreviate to s

is good enough

keep doin

was installed

call up a fri

a

copy and

think of dat
the reportir

ning right on

pass data e
them

can't follow

copy and paste without being extra careful

Grade	Gender	First Name	Last Name
7	M	Freddy	Hanson
8	M	Sandra	Hernandez
7	M	Charlotte	Webster
6	M	John	Johnson
6	M	Michelle	Michelle
7	M	Juan	Paredes
7	M	Janelle	Smith
8	M	Herbert	White
8	M	Snoop	Perro

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

gi
sa
th

ic

nts
born
day

ab

ough

ed

think of data quality as an as-of-date requirement (get everything right on the reporting date)

think of data quality as an as-of-date requirement (get everything right on the reporting date)

pass data entry on to someone who doesn't know the rules or can't follow them

Want data quality? Don't…

Make notes in data fields

Copy and paste from one file (format) to another

be l

ic

r

gi
sa
th

nts
born
day

ab

ough

ed

pass data entry on to someone who doesn't know the rules or can't follow them

call up a friend at IDE and ask for her/him to correct your data

co

"Our student aide will enter everything. Huh? FERPA? "

thin
the

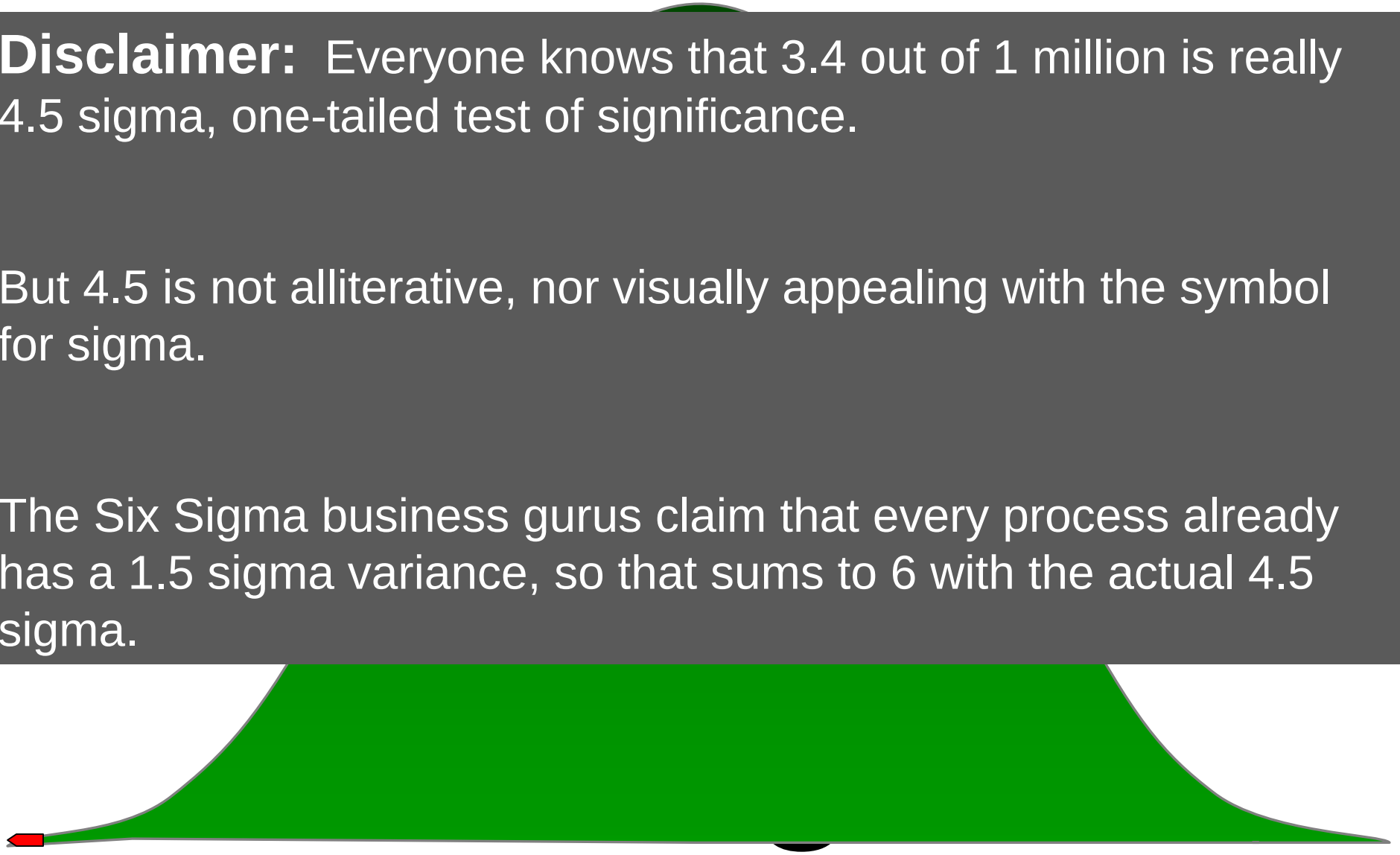
ght on

pass data entry on to someone who doesn't know the rules or can't follow them

Disclaimer: Everyone knows that 3.4 out of 1 million is really 4.5 sigma, one-tailed test of significance.

But 4.5 is not alliterative, nor visually appealing with the symbol for sigma.

The Six Sigma business gurus claim that every process already has a 1.5 sigma variance, so that sums to 6 with the actual 4.5 sigma.



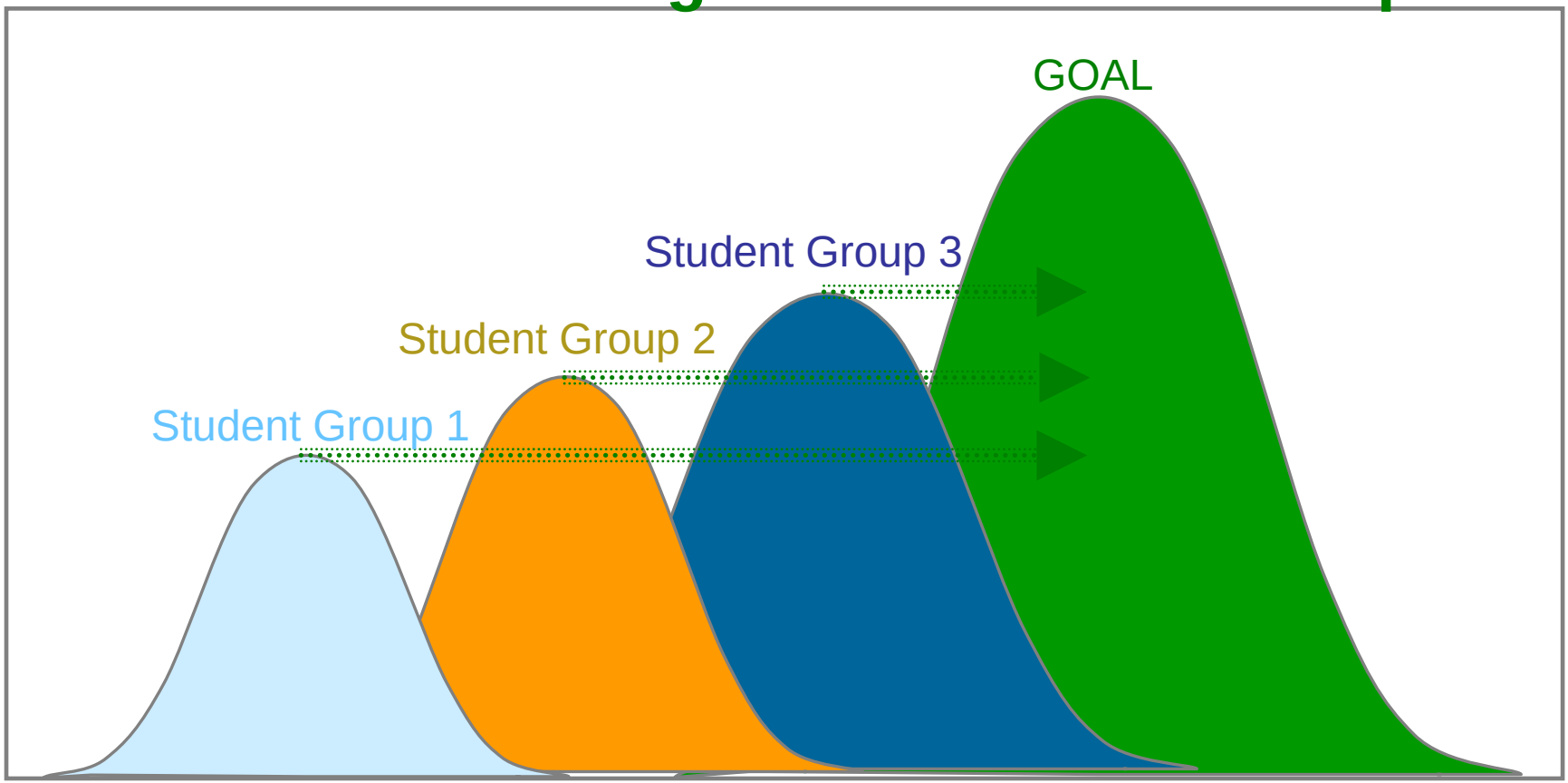
X X X ^

3.4 Defects per Million

ESP Solutions Group

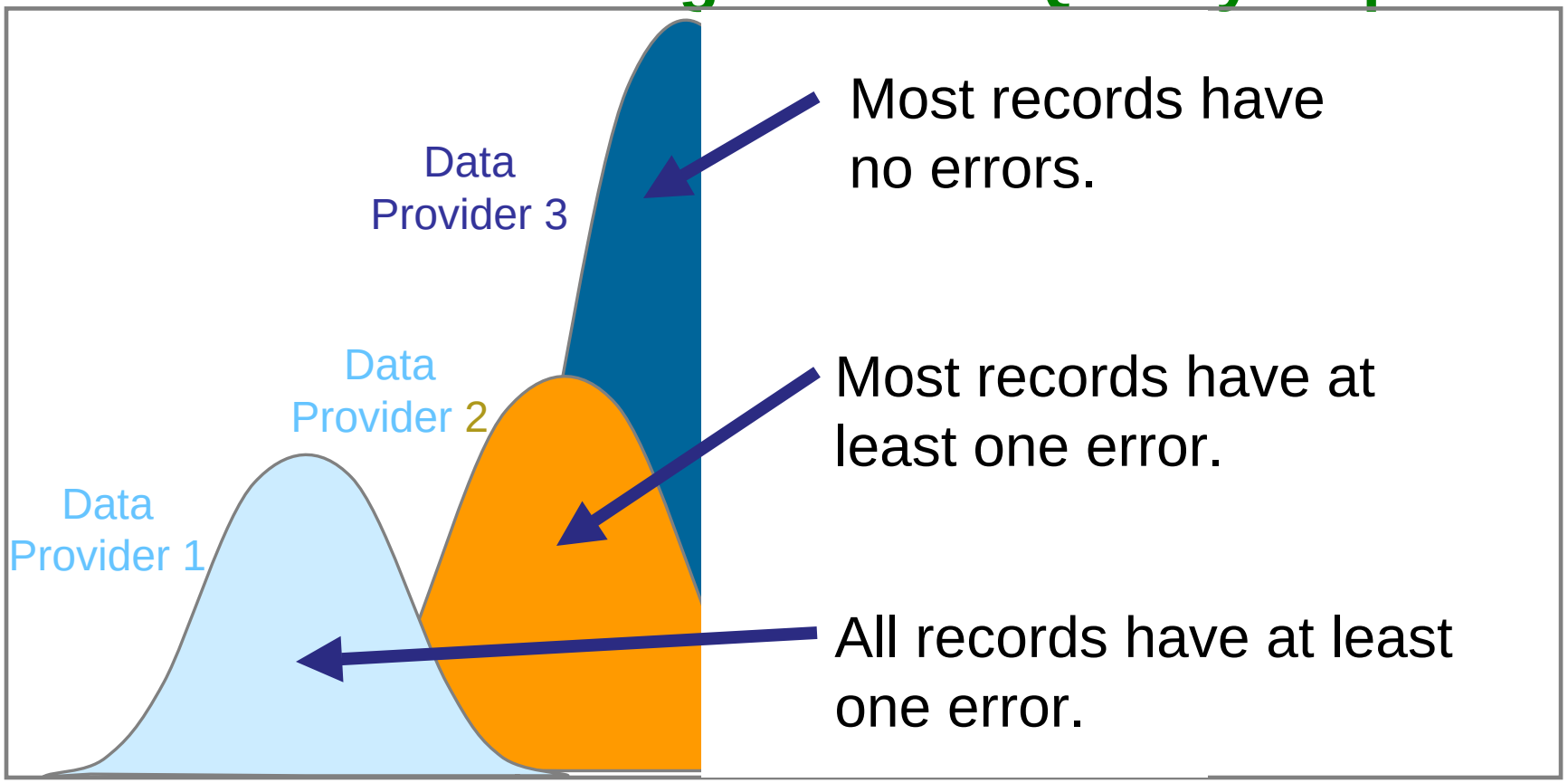
6σ standard allows about 2 Iowa student records to contain an error.

GOAL: Closing the Achievement Gap



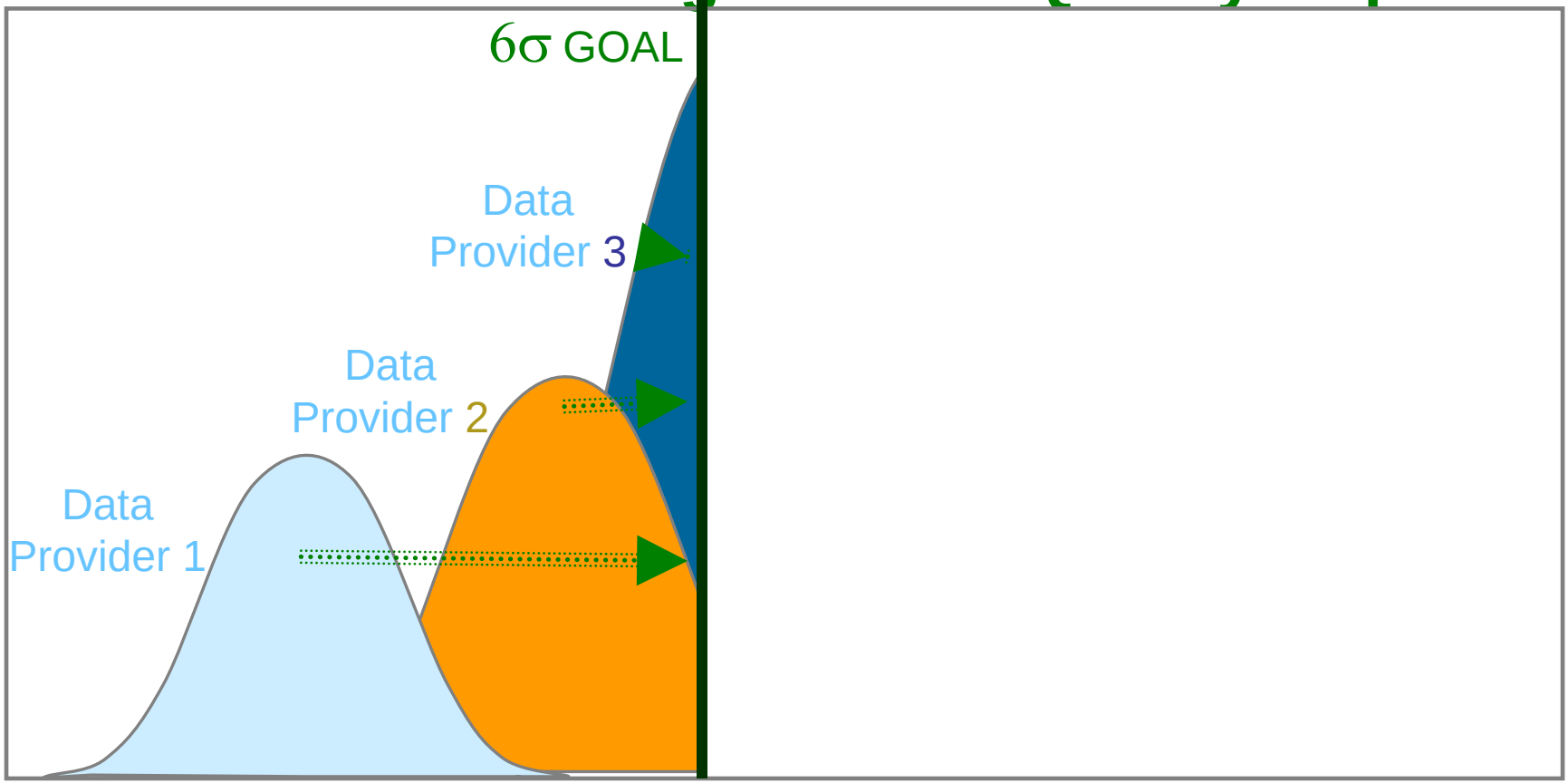
ESP Solutions Group

GOAL: Closing the Data Quality Gap



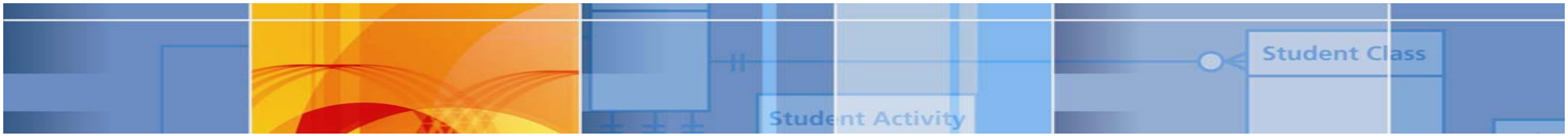
ESP Solutions Group

GOAL: Closing the Data Quality Gap



100% 10% 0% **Fields w/ Errors**

ESP Solutions Group

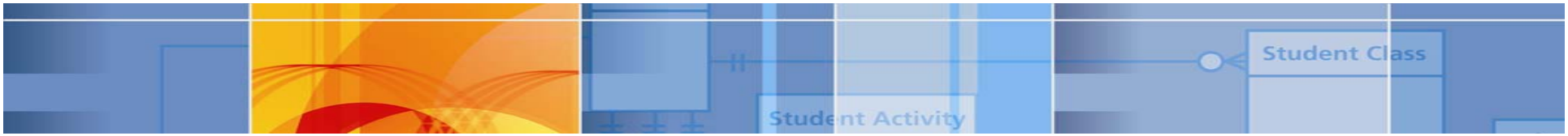


Process Flow of Reported Data:

- **Declaration by the original source of the data (parent)**
- **Entry by the collector**
- **Compiling for reporting**
- **Sending**
- **Receiving**
- **Mapping**
- **Import**
- **Access**
- **Analysis**
- **Formatting**
- **Labeling**
- **Explaining**
- **Interpretation**
- **Use**

Faithfully perpetuate bad data.

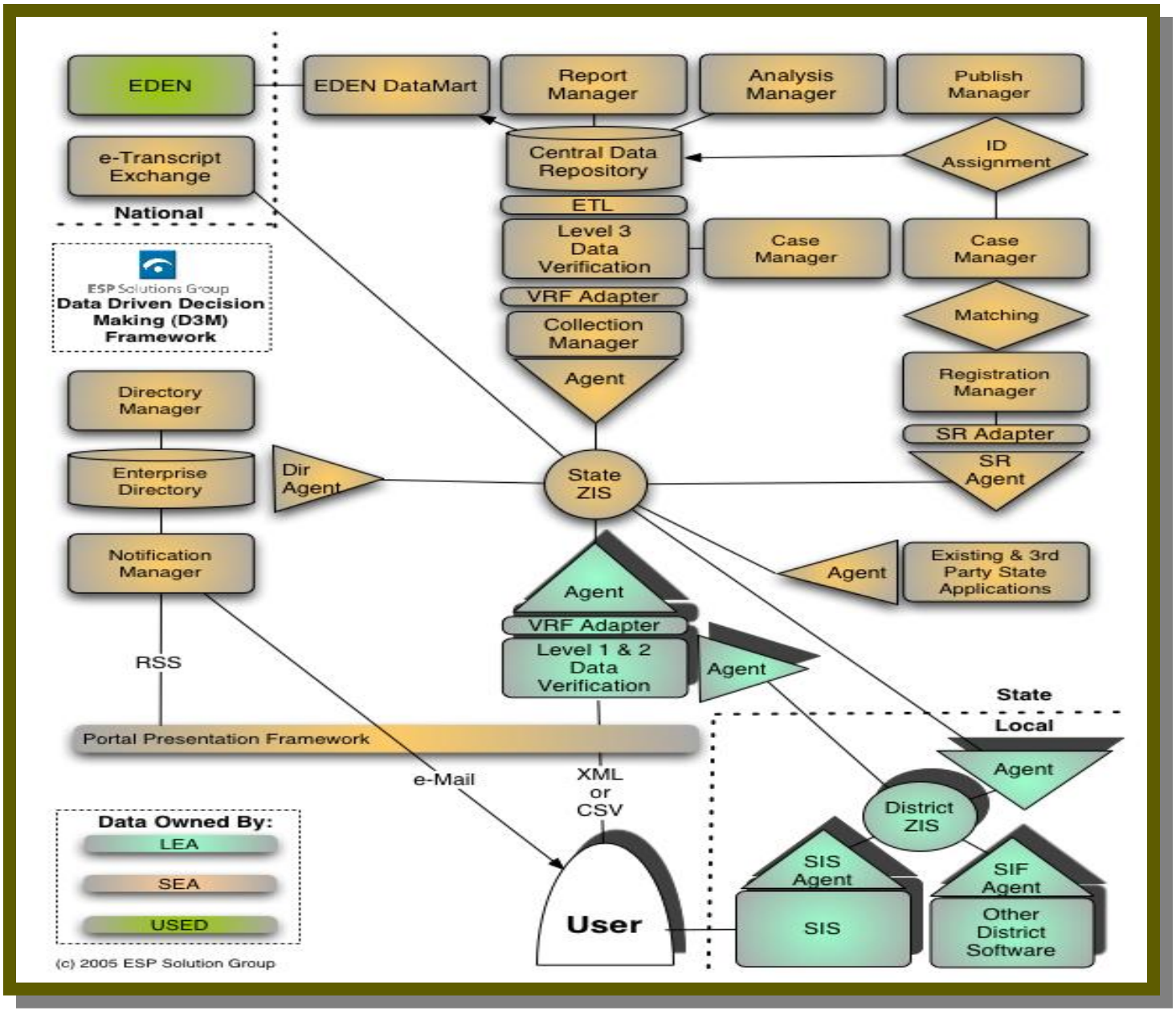
ESP Solutions Group



Other interesting collective nouns:

- **Pride of lions**
- **Exaltation of larks**
- **Battery of tests**
  - **Act of beating or pounding**
  - **Emplacement for artillery**
  - **Cells that generate electricity**
  - **An array of similar things, e.g., tests**

"Kids, tomorrow you will face a battery of tests. They will be lined up to fire questions at you. Any wrong answers will result in an electrical current being zapped through your body."



D3M Framework

Flexible Template for States and Districts

ESP Solutions Group

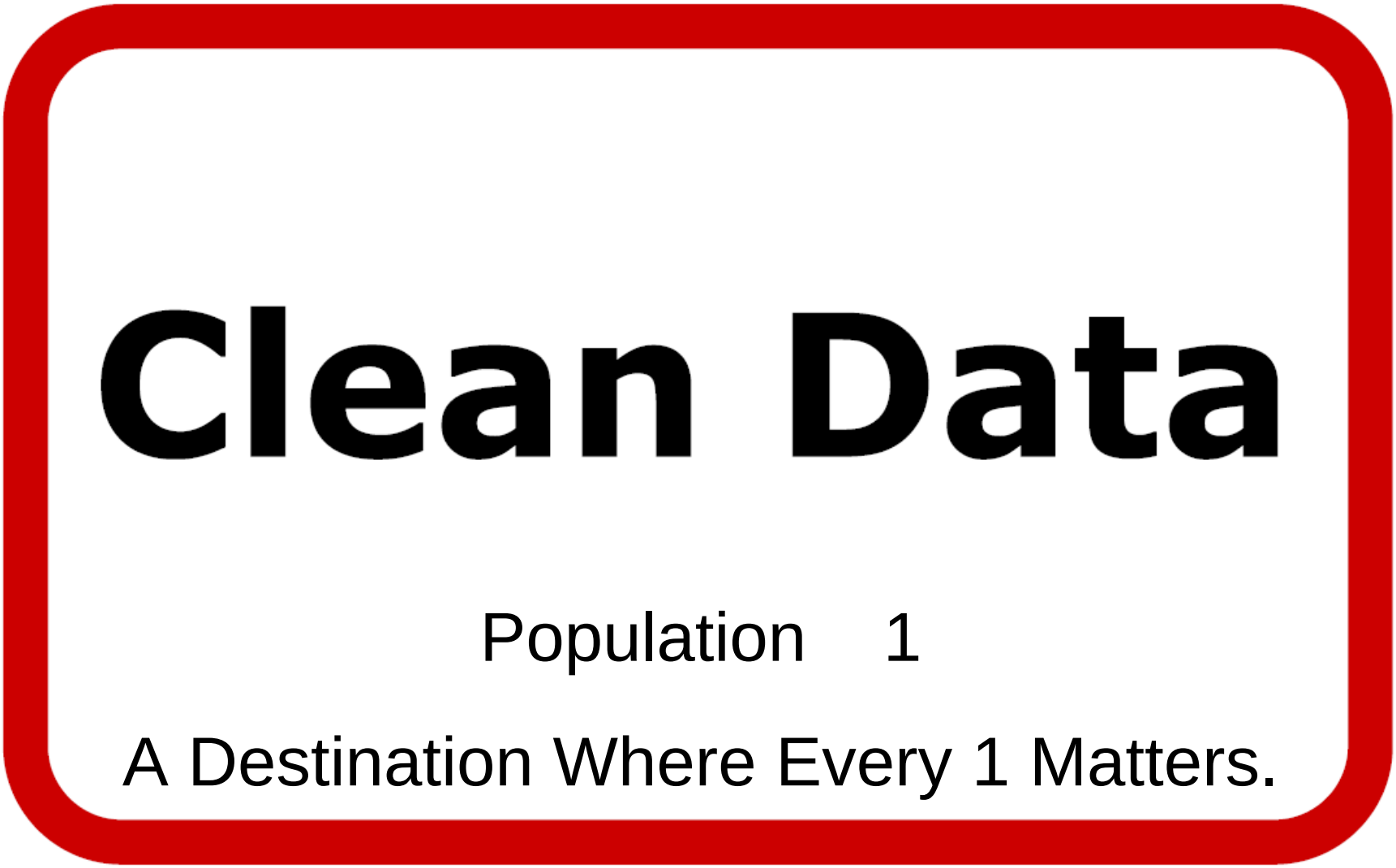
Sign we want posted!

“Clean” is an adjective…
…NOT a verb! NOT a Command!

Adjective…NOT a verb!
Definitely, not a command!

Destination

Quality data are destined for use in data-driven decision making.

Road Sign. A destination. Not a journey!

The purpose for which something is destined.